Computers in Industry 99 (2018) 69–82

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

A visible-range computer-vision system for automated, non-intrusive assessment of the pH value in Thomson oranges

Sajad Sabzi[a], Juan Ignacio Arribas[b,c,*]

[a] Department of Biosystems Engineering, College of Agriculture, University of Mohaghegh Ardabili, Ardabil, Iran
[b] Department of Teoría de la Señal y Comunicaciones e Ingenieria Telematica, University of Valladolid, 47011 Valladolid, Spain
[c] Castilla-León Neuroscience Institute (INCYL), University of Salamanca, 37007 Salamanca, Spain

ARTICLE INFO

Article history:
Received 12 August 2017
Received in revised form 7 March 2018
Accepted 16 March 2018
Available online 25 April 2018

Keywords:
Computer vision
Estimation
Fruit grading
Imperialist competitive algorithm
Neural network
Optimization
pH value
Regression
Response surface methodology
Thomson orange

ABSTRACT

Fruit may be classified for the purposes of usage, packaging and marketing based on the pH (potential of hydrogen) value – a numeric scale used to specify the acidity or basicity of an aqueous solution measured in units of moles per liter of hydrogen ions. In this study, a new approach for the automated and non-intrusive estimation of the pH value of the Thomson navel orange (CRC 969, *Citrus sinensis*) fruit is presented based on visible-range image processing, image feature extraction and with the use of hybrid imperialist competitive algorithm (ICA)-artificial neural network (ANN) regression. Image features studied include length, width, area, eccentricity, perimeter, blue-value, green-value, red-value, width, contrast, texture, roughness and several ratios thereof. Principal component analysis (PCA) is applied to reduce the number of dimensions without loss of important information and a cubic polynomial function of the mean square error (MSE) versus several factors is computed using the response surface methodology (RSM) approach. Results for pH prediction are given and compared with true measured pH values over the entire 100 Thomson orange dataset, including estimated pH scatter regression plots and estimated pH boxplots. Cross validation is performed over 1000 repeated random trial experiments with uniform random train- and test-sample sets (80% training and 20% disjoint test samples). In addition, we provide numerical results based on the levels achieved by response surface methodology (RSM) evaluated over various error coefficients: the sum square error (SSE), the mean absolute error (MAE), the coefficient of determination (R^2), the root mean square error (RMSE), and MSE, resulting in $R^2 = 0.843 \pm 0.043$, $MSE = 0.046 \pm 0.022$, $MAE = 0.166 \pm 0.039$, $SSE = 0.915 \pm 0.425$, and $RMSE = 0.214 \pm 0.146$, over the test set. The results demonstrate that such an automated pH-based sorting system with machine vision using the hybrid ICA-ANN algorithm can accurately compute the pH value of Thomson oranges without any contact with the fruit, and which has clear potential applications in the food industry.

1. Introduction

The estimation of internal quality of products has many applications for example in the food industry to select suitable classification, packing and marketing of fruit. Such determinations currently use intrusive methods with corresponding impacts to the products tested—non-intrusive and automated methods could have significant impacts on productivity, quality and yield. Machine (computer) vision can be applied together with pre-processing algorithms to determine a set of indicators based on visual properties that can predict specific quality features of the product, to be fed in to a high-volume, non-intrusive sorting system.

In the recent years, the interest, usage and importance of machine learning in the food industry has grown rapidly and several relevant papers have been published which compute highly accurate and automatic classification of their respective species under study; Zhao et al. [1] regarding damage recognition on pest-infested oilseed rape leaves, [2] in fish species classification by color, texture and multi-class support vector machine and [3] in fruit classification by neural networks. Tests to measure various physical qualities such as volume, weight, sphericity, pH, taste and sugar content, can be done on different kinds of citrus fruits such as orange [4]. Kondo et al. [5] demonstrated that the quality of orange could be determined by a machine-vision and

* Corresponding author at: Department of Teoría de la Señal y Comunicaciones e Ingenieria Telematica, University of Valladolid, 47011 Valladolid, Spain.
E-mail address: jarribas@tel.uva.es (J.I. Arribas).

https://doi.org/10.1016/j.compind.2018.03.016
0166-3615/

neural network system using different models to estimate at good accuracy pH value and sugar content based on visual and physical features.

An image segmentation method for apple sorting and grading using support vector machines (SVM) and Otsu's method was investigated by [6]. This paper reports on the development of an automatic adjustable algorithm for the segmentation of color images, using linear SVM and Otsu's thresholding method, applied to apple sorting and grading. The method automatically adjusts the classification hyperplane computed by using linear SVM and needs minimum training and time. It also avoids the problems caused by variations in the lighting conditions or the color of the fruit. To evaluate the robustness and accuracy of the proposed segmentation method, tests were conducted for 300 Red Delicious apples using three training input samples with different color characteristics (orange, stripe, and dark red) and their combination.

Iqbal et al. [7] studied predictions of moisture, color and pH value in cooked, pre-sliced turkey hams by a near infrared (NIR) hyperspectral imaging system in the NIR region ({900–1700 nm} wavelength). S(PLS) regression, as a multivariate calibration method to reduce the high dimensionality of the data and to correlate the NIR reflectance spectra with quality attributes of the samples under consideration. Instead of using a wide range of spectra, the number of wavebands was reduced for more stable, comprehensive and faster modelling in the subsequent multispectral imaging system. With the help of PLS regression analysis, nine different wavelengths (927, 944, 1004, 1058, 1108, 1212, 1259, 1362 and 1406 nm) were selected as the optimal wavelengths for moisture prediction, eight different wavelengths (927, 947, 1004, 1071, 1121, 1255, 1312 and 1641 nm) for pH prediction and nine different wavelengths (914, 931, 991, 1115, 1164, 1218, 1282, 1362 and 1638 nm) were identified for color (a*) prediction.

Vazquez-Cruz et al. [8] studied the application of neural networks to estimate carotenoid content during ripening in tomato fruits (*Solanum lycopersicum*). Six tomato ripening stages were considered. The relationship among color parameters (L*, a*, b*, and Hue), maturity stages, and leaf area with the lycopene and β-carotene concentration was analyzed with different regression models. Low R^2 values resulted, showing that lycopene and β-carotene content were not well correlated with color during ripening stages. The objective of this work was to provide a valid ANN model including leaf area index and color readings as inputs to improve the regression models for carotenoid estimations in tomatoes.

There are various image methods which can be potentially used for these purposes; we have opted here for ordinary imaging (visible range camera) due to its simplicity, easy availability of camera equipment, its proven benefits, and appropriateness to be used in an industrial environment. The use of non-destructive methods to predict chemical features in fruits and crops is welcome in agriculture and gardening as internal properties are predicted with no damage done to the products. At the same time, these methods increase the speed of grading and decrease the amount of waste [9].

Several imaging systems include NIR, multispectral and hyperspectral imaging, nuclear magnetic resonance imaging (NMR/MRI), X-ray computed tomography (μCT) [10] and visible range imaging system [11], have been used successfully in non-destructive fields. A visible imaging system is cheaper than other systems and can be used easily. For instance, Mohapatra et al. [11] presented a non-destructive method in banana grading based on changes in the dielectric properties during ripening and image processing. Bananas were stored in a maturity chamber at a temperature of 25 °C. During the ripening process some dielectric properties including capacitance, relative permittivity and impedance were measured. After measuring these properties, one image was captured from each sample using a conventional visible range camera. Three image processing methods were used: Noise Reductant Local Binary Pattern (NRLBP), Local Binary Pattern (LBP) and Completed Local Binary Pattern (CLBP), to predict the degree of fruit maturity. As a final result, two classifiers, Chi-Square distance/nearest-neighbor and Fuzzy-C means (FCM), were successfully used for classification purposes. Cárdenas–Pérez et al. [12], proposed an algorithm to predict Young's modulus during the ripening stages of apple based on color features, firmness and microstructure. Results showed that Young's modulus during ripening stages decreased over time. A significant correlation between Young's modulus, firmness, skin color difference, entropy and fractal dimensions was obtained using Pearson's analysis. As a result, the determination coefficient of Young's modulus predicted from firmness, skin color difference, entropy and fractal dimensions, was 0.95 using multiple linear regression.

Rungpichayapichet et al. [13] studied the prediction of physicochemical features of mango fruits including, firmness, total soluble solids (TSS) and titratable acidity (TA) based on hyperspectral images. Predicted models were developed based on spectral data in visible and infrared ranges using partial least squares regression. Results showed there exists a high correlation between hyperspectral images and firmness, TSS and TA, so that the values of the determination coefficient and root mean squared error between hyperspectral images and firmness, TSS and TA, were 0.81, 2.83 and 0.5 (determination coefficients), and 2, 0.81 and 0.24 (root mean square errors), respectively.

As noted above, researchers have focused in recent years on the prediction of physicochemical properties based on color features, and spectral data. Further development of these non-destructive methods to non-intrusively evaluate physicochemical properties in the different stages of fruit and crop ripening may allow automated management, harvesting, handling, post harvesting and warehousing to become a reality. Based on previous analyses, we focused in this study on the most difficult and yet cheapest non-destructive image prediction system, the visible range imaging system (ordinary camera).

The specific aim of this study is to predict the pH value of Thomson orange based on color and texture features from ordinary images of fruit peel (exterior) using simple yet very effective image processing techniques together with the ICA-ANN methodology. Such a proposed methodology is novel and to our best knowledge there are no similar reports published in the literature so far.

Fig. 1 shows a system block diagram of the proposed system for automatic and non-intrusive computer vision sorting of Thomson

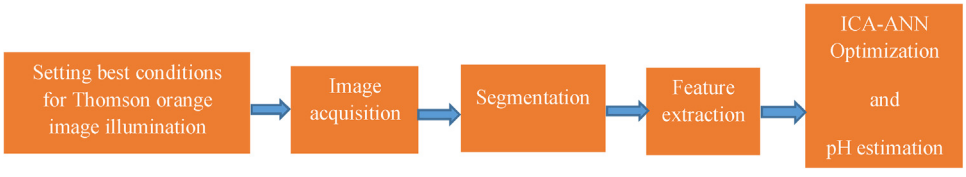

Fig. 1. A system block diagram for the automatic non-intrusive computer vision sorting of Thomson orange based on pH value.